第 65 卷 第 6 期　　天 文 学 报　　Vol. 65 No. 6
2024 年 11 月　　ACTA ASTRONOMICA SINICA　　Nov., 2024
doi: 10.15940/j.cnki.0001-5245.2024.06.006

基于CSST的暂现源探测方法*

赵杰明[1†]　王晓锋[1,2]　李国亮[3]　张天萌[4]
(1 清华大学物理系 北京 100084)
(2 北京天文馆 北京 100044)
(3 中国科学院紫金山天文台 南京 210023)
(4 中国科学院国家天文台 北京 100012)

摘要　中国巡天空间望远镜(China Space Survey Telescope, CSST)是中国载人航天工程规划建设的以大规模天文巡天为主任务的2 m口径空间光学望远镜, 兼具高空间分辨率与大视场特性, 性能优异. 暂现源是CSST的重要科学目标之一. CSST作为空间望远镜, 与地面望远镜的图像噪声组成有着明显不同. 因此, 不能直接套用先前地基巡天项目的暂现源探测方案, 需要针对性开发专属的探测方法. CSST的暂现源探测方法以CSST仿真数据为基础进行开发、验证, 旨在尽可能准确地识别CSST主巡天数据中的暂现源. 方法以图像相减法为基础, 首先获得观测图像所对应残差图像, 并利用滤波方法给出残差图像泊松噪声分布; 然后使用渐变阈值间接统一背景噪声水平, 对残差图像进行目标源测光得到候选源列表; 最后, 通过判断候选源与当地泊松噪声的偏差对真实暂现源进行筛选, 结合原始图像信息与观测图像测光分类信息后, 进一步对候选源列表的假源进行剔除, 最终输出暂现源星表. 方法在共计20000个暂现源的多轮仿真测试中, 能够以平均95.9%的准确率对观测图像中的暂现源进行筛选. 相较于先前工作, 还更加全面且定量地给出了测试所得的不同亮度暂现源所对应的探测率. 测试结论验证了CSST暂现源探测方法的可行性、泛化能力与稳定性. 该方法也为空间望远镜暂现源探测任务提供了理论与编程基础.

关键词　方法: 数据分析, 技术: 图像处理, 望远镜, 巡天
中图分类号: P141;　**文献标识码**: A

1　引言

1.1　科学背景

暂现源是一类短时间内亮度急剧增加而后降低甚至消失的非周期性光变天体. 我们所熟知的超新星就是一类典型的暂现源. 随着多波段、多信使天文观测时代的到来, 伽马射线暴、引力波事件等光学辐射对应体也成为了重要的暂现源. 这些暂现源通常伴随着剧烈能量释放与极端物理条件, 其中大部分都可通过多波段、多信使进行协同观测, 具有很高的科研价值, 是时域天文学的重要研究对象. 以Ia型超新星为例, 获得大样本的Ia型超新星不仅可以直接得到其宿主星系的精确距离定标, 还可以降低Ia型超新星的演化误差, 利用二者共同限制宇宙学参数的测量精度. 再以引力波事件光学辐射对应体为例, 同样利用大样本进行统计分类, 才能做到观测数据和模型的相互验证, 进一步理解产生极端相对论喷流、千新星以及引力波辐射的

2023-08-22收到原稿, 2023-12-13收到修改稿
*中国载人航天工程巡天空间望远镜专项科学研究项目(CMS-CSST-2021-A12)资助
†maoxiuxingtuanm45@163.com

物理机制. 同时, 利用大样本的引力波标准汽笛来测量基本宇宙学参数, 也有望解决目前宇宙学参数测量中的不自洽问题. 现阶段时域天文学的研究亟需发现更多的暂现源以建立规模可观的样本库. 因此, 寻找、观测暂现源即成为了中国巡天空间望远镜(China Space Survey Telescope, CSST)的重要科学目标之一[1-2]. 理论预研工作表明, 在不借助强引力透镜效应的情况下, CSST有能力探测到最高红移为2的Ia型超新星[3]. 望远镜10 yr在轨期间对10 deg^2天区的观测, 有望探测到1800颗左右$1.0 < z < 1.3$的高红移Ia型超新星, 利用这些样本同样可以显著减少当前宇宙学参数的测量误差[1].

1.2 CSST望远镜

CSST是中国载人航天工程规划建设, 以大规模天文巡天为主任务的2 m口径空间光学望远镜. 其搭载了巡天模块、多通道成像仪、积分视场光谱仪等多个科学载荷, 同时具备近紫外-可见光-近红外多波段观测能力. 望远镜主光机采用了28 m焦距的离轴三反光学系统. 离轴三反光学系统的优势在于其消除了折射色差, 同时兼具大视场并保证了光学镜组之间没有遮挡. 因此, 相对于光路中存在镜架支撑结构遮挡的光学系统具有更好的成像质量. 望远镜所搭载的巡天模块主体由一台30片探测器拼接而成的大型焦面相机组成, 主焦面30片探测器80%的有效观测视场共约$1.1^\circ \times 1^\circ$. 主光学系统角分辨率约为$0.13''$, 每片探测器每像元对应视角约为$0.074''$. CSST主巡天模式(常规巡天模式)采用对每个天区由一片探测器曝光一次的观测策略. 数据处理流水线将上次观测同一天区的观测图像作为模板图像, 与本次观测图像一同输入暂现源探测模块进行暂现源探测. 主巡天模式下, 每天拍摄约300次, 每片探测器每次曝光时间150 s, 读出时间约40 s, 观测同一天区的时间间隔不定[4].

对于给定的光学系统与外部环境, 光学系统PSF (Point Spread Function)的FWHM (Full Width Half Maxima)与R_{EE80} (80%能量集中度半径)是定量表征成像解析能力与成像质量的重要数值特征. 为直观说明CSST的成像特点, 图1对比展示了CSST的PSF轮廓(PSF轮廓数据来自CSST主巡天仿真数据产品Cycle3)与斯隆数字巡天(Sloan Digital Sky Survey, SDSS)[2]及哈勃空间望远镜(Hubble Space Telescope, HST)的PSF轮廓[5]及其截面.

从图1可以看到, CSST的PSF非常尖锐. 与作为地面光学望远镜的SDSS相比, CSST作为空间光学望远镜无论是代表成像解析能力的FWHM, 还是代表成像质量的R_{EE80}都更为优异, 因为后者受到大气层视宁度(Seeing)影响且PSF还会随时间不断变化. R_{EE80}代表的成像质量会影响望远镜对近邻, 特别是密集星场的天体分辨能力. 近邻天体可能会因为PSF的外围延展而使得轮廓模糊相接而无法区分(Blending效应), Blending同时会造成对天体的测光偏差. 事实上, 经过模拟[3], CSST的R_{EE80}是同时期大型综合巡天望远镜(Large Synoptic Survey Telescope, LSST)的约1/4. 可见CSST作为空间望远镜, 在不受大气层影响的情况下相对于地基望远镜在成像质量上有着无法比拟的优势. 同为空间光学望远镜的HST与CSST相比, 二者均有着锐利且几乎不随时间变化的PSF. 从图1 (d)可以看到, HST的PSF有着更小的FWHM, 中心峰更加尖锐, 这说明HST相对于CSST有着更强的解析能力. 但HST的成像系统受到副镜支架遮挡与衍射的影响, 使得PSF外围更加延展, 导致其成像质量不如CSST, 最终使得二者的R_{EE80}处于相同水平.

[1]《空间站多功能光学设施科学白皮书》查阅网址: https://nao.cas.cn/csst/
[2]《空间站多功能光学设施科学白皮书》5.4. 其他科学. 空间站多功能光学设施科学团队. 2020-03
[3]《空间站多功能光学设施科学白皮书》5.1. 宇宙学. 空间站多功能光学设施科学团队. 2020-03
[4]《空间站多功能光学设施科学白皮书》1.1. 设备基本情况. 空间站多功能光学设施科学团队. 2020-03
[5]查阅网址: https://www.stsci.edu/hst/instrumentation/

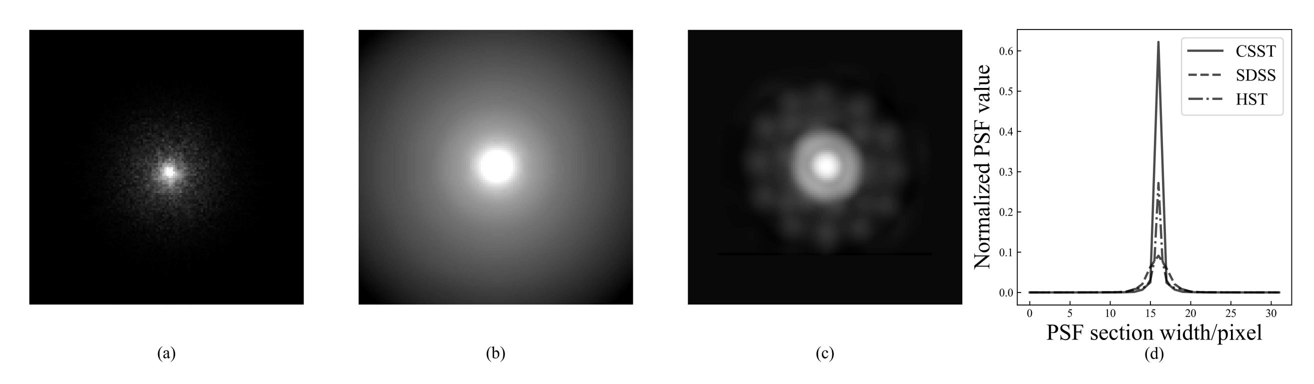


图 1 CSST望远镜主巡天i波段成像PSF与其他望远镜i波段PSF对比. (a) CSST望远镜主巡天成像i波段PSF模拟轮廓; (b) SDSS巡天成像i波段PSF经验轮廓; (c) HST-WFC3相机i波段PSF经验轮廓, 以上轮廓大小均为32 × 32像素, 且均经过灰度值对数处理以便直观展示PSF之间的差异; (d) 3望远镜PSF截面对比.

Fig. 1 Comparison of PSF in i-band between CSST and other telescopes. (a) Simulated PSF profile in i-band of CSST main survey imaging. (b) Empirical PSF profile in i-band of SDSS survey imaging. (c) Empirical PSF profile in i-band of HST-WFC3 imaging, the above profile sizes are all 32 × 32 pixels, all processed with logarithmic grayscale values to visually display the differences between PSFs. (d) Comparison of PSF cross-sections from three telescopes above

综上所述, CSST同时拥有大视场、高空间分辨率、高成像质量以及较短曝光时间的优秀巡天观测硬件条件. 作为CSST的科学目标之一, 无论是超新星还是其他爆发现象的光学辐射对应体, 这些暂现源通常有着很强的星系背景亮度. 源自身与宿主星系的光度轮廓会有很大概率出现重叠. CSST的高空间分辨率与高成像质量使其有能力分辨并解构(Deblending)出这些对地面光学望远镜而言, 因PSF制约而淹没在宿主星系星光背景中的暗源. 大视场及较短曝光时间则大大提高了其巡天效率, 这些优势使得CSST有能力在观测中发现更多、更暗的暂现源.

1.3 暂现源探测模块

暂现源探测模块是CSST数据处理流水线中的一个模块, 旨在尽可能准确地识别CSST主巡天数据中的暂现源, 为后续时域天文学研究提供大量新样本. 由于CSST还没有实际投入运行, 为保证其在轨运行后科学产出的及时性、可靠性, CSST科学团队需要提前对数据处理流水线的各个模块进行理论预研、开发及测试验证. 本工作即基于CSST的图像特点, 针对性地进行了暂现源探测方法的理论预研与论证. 同时利用Python3编程进行模块开发, 将所有流程方法在不使用第三方软件的条件下予以实现. 继而, 使用CSST仿真数据对模块进行测试, 以验证流程与方法的可行性, 迭代式地通过测试中所暴露的问题对流程与方法进行改进. 在模块开发完成后, 将最终进行多轮独立测试以获得模块的性能指标.

1.4 CSST仿真程序

CSST科学团队为望远镜性能评估、数据处理流水线测试以及相关科研课题理论预研开发了CSST主巡天仿真程序. 仿真程序按照望远镜真实成像原理生成主巡天仿真图像, 程序考虑到了常数弱透镜剪切、宇宙线、背景天光、宇宙学效应等物理效应以及点扩散、读出噪声、平场渐晕、快门效应、暗电流、电荷转移低效(Charge Transfer Inefficiency, CTI)、坏像列、失效像元、电子弥散、电荷耦合元件(Charge-Coupled Device, CCD)饱和溢出等仪器效应的影响, 相应的参数设置尽可能做到了与实际预期相符[6–7]. 仿真程序最终能够给出单片探测器曝光得到的各波段主巡天FITS图像, 对应尺寸约9.2 k×9.2 k像素[8]. 本工作所使用的仿

[6]《CSST主巡天仿真数据数据产品说明》查阅网址: http://www.bao.ac.cn/csst/research/software/
[7]《CSST主巡天仿真数据数据产品说明》1.1综述. 李楠, 张鑫, 刘德子, 等. 2021-05-12
[8]《CSST主巡天仿真数据数据产品说明》2.5仪器效应, 3仿真结果. 李楠, 张鑫, 刘德子, 等. 2021-05-12

真数据来自CSST主巡天仿真数据Cycle 3.

1.5 与现有解决方案的差异

CSST暂现源探测方法及模块开发与现有的成熟解决方案[4–7]在先决条件上有所不同, 导致其不能直接套用这些方案. 需要以图像相减法(图像差异分析, Difference Image Analysis, DIA)[8]为核心思想重新设计整体流程并开发新方法以解决差异之处. 其与已有方案主要差异是: 暂现源探测模块着重探测以超新星或其他爆发事件光学辐射对应体为主要目标的暂现源, 这类暂现源一般有着来自宿主星系较大的背景亮度. 由于CSST有着高空间分辨率与高成像质量, 在残差图像中进行真实暂现源、变源与假源区分时, CSST残差图像中的假源主要来自于星光背景噪声涨落产生的残影假源以及少数无法与点源区分的宇宙线. 相比之下, 地面望远镜残差图像中的假源则主要来自于因大气条件变化所产生的观测图像与模板图像之间的卷积校准异常. 为解决该差异点, 本工作重新推导建立了针对CSST图像的噪声模型, 并创新性地开发了残差图像判定方法. 方法使用渐变阈值间接统一残差图像背景噪声水平后, 对残差图像进行测光获得候选源列表, 通过考察候选源列表中各个源与当地泊松噪声背景的差异来进行真实暂现源筛选.

2 暂现源探测整体流程设计

本工作采用来自CSST主巡天仿真数据Cycle 3的标准i波段FITS (Flexible Image Transport System)图像. 在原始观测图像进入暂现源探测模块前, 已经经过了上游流水线的数据预处理, 包括: 探测器本底偏置校准、平场改正、噪声均值测定、坏点坏列标定、饱和像素标定、像场畸变修正、坐标校准、流量定标、图像重采样对齐等步骤. 模板图像作为望远镜对同一天区先前的观测图像, 也经过了同样的预处理流程. 因此, 模块的输入为: 模板图像以及与模板图像具有相同视场但观测时间不同, 并几乎完全对齐完成流量修正的观测图像. 模块的最终输出是暂现源星表.

暂现源探测模块是以图像相减法为核心思想而重新设计的解决方案, 模块分为3大部分: 噪声分析与背景处理、测光分类以及残差图像识别. 噪声分析与背景处理以及测光分类两大部分所涉及方法均属于现有成熟方案. 本工作在其基础上进行了针对性的改进与优化, 因此对前两部分只在下文中做简单功能介绍, 不详细展开说明具体方法, 主要展示重要结论. 残差图像识别部分作为CSST暂现源探测模块的核心方法与创新部分, 将在正文中详细介绍其具体方法与测试结论. 在进行方法介绍之前, 首先展示暂现源探测模块流程图. 框图中说明了每部分具体功能与数据流, 具体流程如图2所示.

2.1 背景噪声分析

模块第一部分主要作用是利用滤波方法对输入图像进行噪声分析以去除天光背景, 同时建立噪声模型并给出图像统计量, 为宇宙线去除及后续方法提供阈值参考. 在CSST图像中, 占主导的噪声为符合泊松分布的光子散粒噪声(Shot noise)[9]. 除此之外还需考察暗电流热噪声、探测器读出噪声与数字化噪声. 暗电流热噪声同样符合泊松分布, 读出噪声与数字化噪声则符合正态分布[10]. 读出噪声与数字化噪声均与信号无关, 这种与输入信号无关且与输入信号呈相加关系的噪声被称为加性噪声(Additive noise), 可以将二者合并记其标准差为σ_{add}. 对给定的观测图像, 设g为各通道校准后单张图像增益(Gain)的增益系数. 定义为图像单位灰度值(减去偏置后)所代表的实际电子数, 且已知$g = 1.1$. 设图像曝光时间为t, 同时已知$t = 150\,\mathrm{s}$[9].设探测器CCD的量子效率为α, 定义为光子转化为电子的效率. 设曝光累计光通量为Φ以及给定温度下热电子密度为n, 设暂现源探测模块的输入图像为I_{input}, 且原始图像在经过预处理后通过暗场积分测定去除了暗电流热噪声均值, 也通过仪器实验标定去除了读出噪声与数字化噪声均值.

[9] 《CSST主巡天仿真数据数据产品说明》2.5仪器效应, 3仿真结果. 李楠, 张鑫, 刘德子, 等. 2021-05-12

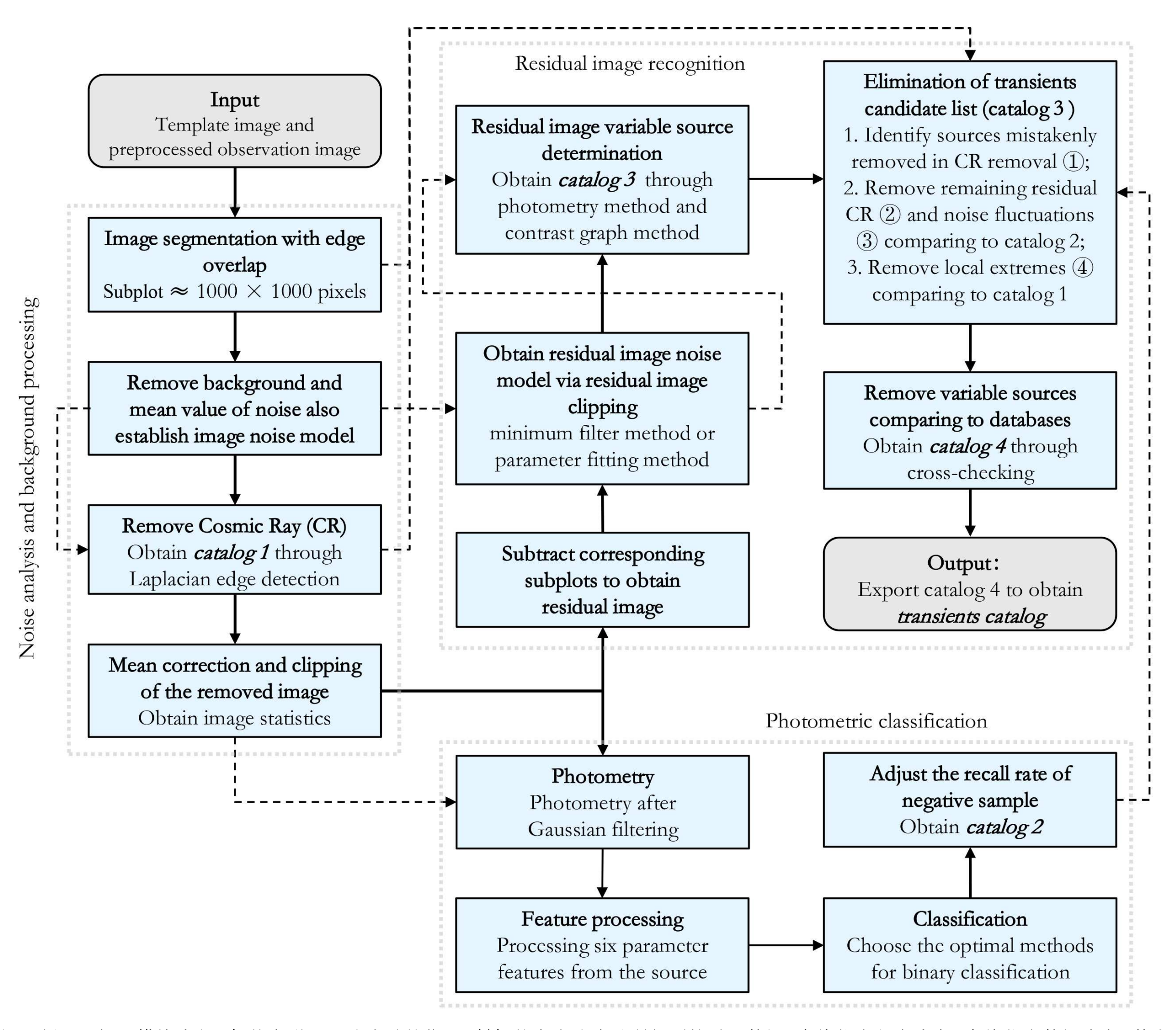


图 2 暂现源探测模块流程. 粗体字说明了该步骤的作用, 斜粗体字为该步骤所得到的重要数据, 实线代表程序流向, 虚线代表数据流向; 其中星表1为宇宙线标记星表, 星表2为测光分类星表, 星表3为暂现源候选列表, 星表4为暂现源星表.

Fig. 2 The process of transient detection module. While bold text explains function of the step, italic bold text represents important data of the step, solid line represents program flow, dashed line represents data flow; Among them, catalog 1 is a cosmic ray marker catalog, catalog 2 is a photometric classification catalog, catalog 3 is a transient candidate list, and catalog 4 is a transient catalog.

将符合泊松分布的数据记为Poisson(·), 符合正态分布的数据记为Normal(·), 记求期望为$E(\cdot)$. 有推论: 图像噪声的方差$\sigma^2(I_{\rm input})$与图像期望$E(I_{\rm input})$呈线性关系:

$$\begin{aligned} E(I_{\rm input}) &= \alpha\Phi tg \\ \sigma^2(I_{\rm input}) &= (\alpha\Phi + n)tg^2 + \sigma^2_{\rm add} \\ &= E(I_{\rm input})g + ntg^2 + \sigma^2_{\rm add} \ . \end{aligned} \tag{1}$$

记λ为泊松分布参数, 在光通量足够大即泊松分布期望足够大时, 泊松分布可以用正态分布近似:

$$\begin{aligned} p(X=k) = \frac{\lambda^k}{k!}{\rm e}^{-\lambda} \to \frac{1}{\sqrt{2\pi\lambda}}{\rm e}^{-\frac{(k-\lambda)^2}{2\lambda}}\ , \\ {\rm Poisson}\ (\lambda) \sim {\rm Normal}(\lambda,\lambda)\ . \end{aligned} \tag{2}$$

对于CSST图像, 其像素灰度值与该像素光子计数

之间的关系是:

$$像素接收光子数 = 像素灰度值 \times \frac{g}{\alpha} \times t\,. \quad (3)$$

经过后续基于中值滤波的图像像素值统计, 典型CSST主巡天图像150 s曝光的天光背景均值, 约对应24.6个光子. 因此, 在绝大多数输入图像中, 可以认为天光背景像素值更加接近泊松分布, 而较为明亮的星系背景则会更加接近正态分布.

理想情况下, 暗电流热噪声的均值在暗场修正中被完全去除, 读出噪声与数模转化噪声在去除标定值后均值为零. 在实际情况中, 这些噪声的均值可能经过测定修正后显著减小. 但此时, 这些噪声的残留均值远小于天光背景均值. 因此, 不妨在此后推导过程中把这些真实情况中可能残留的各种噪声均值全部算入天光背景均值之中. 由(1)式可知, 在曝光时间给定的情况下, 暗电流热噪声的方差与前述加性噪声的方差一样为常数, 故将二者合并为常值噪声, 并定义其标准差为$\sigma_{\rm const}$. 常值噪声相对天光背景噪声是小量, 可忽略不计, 具体数值可通过暗场曝光确定.

根据以上噪声分析的结论, 可以将输入图像分为标准差为0的纯均值与均值为0的纯噪声, 这些噪声均无法被去除, 也就是: 输入图像可表示为天光背景均值、科学数据均值、天光背景噪声、科学数据噪声以及常值噪声之和.

2.2 测光分类

模块第二部分主要作用是对去除天光背景后的观测图像, 应用DAOPHOT (Dominion Astrophysical Observatory PHOTometry)中的DAOStarFinder算法[11]测光, 并利用监督学习对测光源进行二分类. 二分类以该源是否为科学数据为标签, 标签中的科学数据类在此特指输入星表中的所有3类天体: 恒星、星系与活动星系核(Active Galactic Nucleus, AGN), 而非科学数据类在此则指探测到如宇宙线残留、明亮背景噪声涨落等原因所产生的各种假源. 处理后的观测图像测光结果统计如表1所示.

表1　输入星表与测光列表中的天体占比
Table 1 Proportion of celestial bodies in input catalog and photometry catalog

Category	Star	Galaxy	AGN	Fake
Input catalog	0.01	0.88	0.08	/
Photometry catalog (Fake excluded)	0.06	0.91	0.03	/
Photometry catalog (Fake included)	0.05	0.8	0.03	0.12

观察表1可以发现, 星系、AGN在测光源列表中相对占比降低, 而恒星相对占比提高. 这是因为恒星分布相对亮度较高, 淹没在天光背景中的部分占比较小. 在测光同时, 通过模型拟合可以获得每个源所对应模型的6个数值参数: 孔径流量、半光面积、离心率、峰值、半长轴、半短轴, 利用这6项参数作为分类特征进行分类器选择. 本工作将每种备选模型都进行了两分类与四分类(作为参考)的训练. 并从10折交叉验证的平均受试者工作特性曲线下面积(Area Under Curve, AUC)[12]、概率校准曲线[13]、模型扩展性、学习曲线以及训练时间成本5个维度来综合评价分类器性能. 经过基于分类器原理在此任务中所具有的可解释性为判别标准的初步筛选后, 决定考察核函数支持向量机(Kernel Support Vector Machine, Kernel SVM)[14]、多层感知机(Multilayer Perceptron, MLP)[15]、随机森林(Random Forest, RF)[16]、自适应增强(Adaptive Boosting, AdaBoost)[17]、梯度提升决策树(Gradient Boosting Decision Tree, GBDT)[18]、极端梯度提升(eXtreme Gradient Boosting, XGBoost)[19]以及高斯过程分类器(Gaussian Process Classifier, GPC)[20]这7种监督学习模型. 所有模型的训练样本数量都以测试集正确率向训练集正确率收敛为最小样本判断标准, 以避免出现过拟合与欠拟合的现象. 所有模型的参数调整均以10折交叉验证正确率最佳为判断标准进行参数调整. 在遇到需要多参数联合调整的模型时, 首先基于模型原理尽量分离能够单独调整的参数, 对于没有明确前后调整顺序的参数则以交叉迭

代的方式进行多参数调整. 经过调整与测试, 各分类器在本工作任务中的性能如表2所示.

表 2 测光源列表分类任务中各分类器性能
Table 2 Classifier performance on photometry catalog classification

Classifier	Avg Accuracy	Two CAT				AUC	Four CAT
		True Positive	False Positive	False Negative	True Negative		Avg Accuracy
Kernel SVM	0.912	1964	162	24	101	**0.88±0.04**	0.867
MLP	0.891	1951	137	37	126	**0.89±0.03**	0.852
RF	0.902	1985	19	3	244	**0.88±0.04**	0.858
AdaBoost	0.898	1890	131	98	132	**0.89±0.03**	0.847
GBDT	0.904	1899	133	89	130	**0.90±0.03**	0.838
XGBoost	0.941	1980	95	45	131	**0.94±0.03**	0.904
GPC	0.933	1956	136	32	127	**0.92±0.02**	0.903

通过表2可以得出结论, 极端梯度提升树、高斯过程分类器与梯度提升树3者有着最好的分类正确率及泛化能力. 由于稳定性、模块兼容导致的二次可操作性原因, 工作暂时放弃使用极端梯度提升树, 而选择了利用梯度提升树或高斯过程分类器对测光源列表进行分类. 在对分类器进行偏向假源查全率的调整后, 获得最终分类结果. 这是因为, 如若分类器将真实源判断为假源并遮罩或标记, 可以通过对比两张原始图像的最小值滤波图找回误去除的真实源. 但若分类器将假源判断为真实源, 则在残差图像中将出现该假源产生的暂现源, 从而影响最终的暂现源探测率与准确率.

3 CSST暂现源探测方法

3.1 残差图像

经模块第一部分处理后的输入图像与模板图像进行对应像素相减, 即得到残差图像. 首先考察残差图像的组成, 由2.1节结论及独立随机变量线性组合的性质可以进一步得出以下结论: 残差图像可表示为科学数据噪声、天光背景噪声与常值噪声之和. 经第一部分处理后的观测图像与模板图像, 理论上二者的科学数据均值没有区别. 除暂现源、变源外其余均值应相减为零. 3种可以近似为正态分布的非负噪声在削波后都符合半正态分布. 由独立随机变量的性质可知, 独立随机变量X、Y的线性组合满足: $E\left(aX \pm bY\right) = aE\left(X\right) \pm bE\left(Y\right), \sigma^2(aX \pm bY) = a^2\sigma^2(X) + b^2\sigma^2(Y)$, 其中$a$、$b$表示线性组合的系数, σ^2表示随机变量的方差. 因此, 设半正态分布相对应的正态分布为Normal $(0, \sigma^2)$, 并记二者方差之间的系数为$\varepsilon = 1 - \frac{2}{\pi}$, 则半正态分布的均值$\mu_{\rm h}$与方差$\sigma_{\rm h}^2$为:

$$\mu_{\rm h} = \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \quad, \sigma_{\rm h}^2 = \left(1 - \frac{2}{\pi}\right)\sigma^2 = \varepsilon\sigma^2 \,. \tag{4}$$

继而可以写出残差图像的期望与方差:

$$\begin{aligned} E(I_{\rm residual}) &= 0\,, \\ \sigma^2(I_{\rm residual}) &= 2\varepsilon\alpha\Phi t g^2 + 2\varepsilon\sigma_{\rm const}^2 \,. \end{aligned} \tag{5}$$

残差图中像素值的分布为两符合半正态分布随机变量之差, 其符合准三角分布而并非正态分布. 但准三角分布仅在接近分布均值附近与正态分布有差异, 且差异随方差增大而减小. 因此可以用正态分布对其进行近似.

3.2 残差图像噪声模型

根据(1)式, 图像噪声的方差与图像期望呈线性关系, 且实际图像的期望(实际应用中用平均估计代替求期望, 记为$<\cdot>$)与光通量成正比:

$$E\left(I_{\text{input}}\right)=\alpha\Phi tg\,,\qquad \sigma^{2}\left(I_{\text{input}}\right)=E\left(I_{\text{input}}\right)g+\sigma_{\text{const}}^{2}\,. \tag{6}$$

因此在应用中, 可以通过图像像素均值反推当地噪声标准差, 如下所示:

$$\sigma_{\text{residual}}=\sqrt{2\varepsilon\alpha\Phi tg^{2}+2\varepsilon\sigma_{\text{const}}^{2}}=\sqrt{2\varepsilon g<I_{\text{input}}>+2\varepsilon\sigma_{\text{const}}^{2}}\,. \tag{7}$$

而图像像素均值则可以通过滤波方法获得. 记滤波核尺寸为$m\times m$的均值滤波器为$\text{Mean}_{m\times m}$, 记滤波核尺寸为$m\times m$的中值滤波器为$\text{Median}_{m\times m}$, 将卷积符号记为$\circ$. 如若所考察滤波核内像素均值与中位数的差值小于等于0.3倍像素标准差, 则所考察区域内像素的平均估计取为:

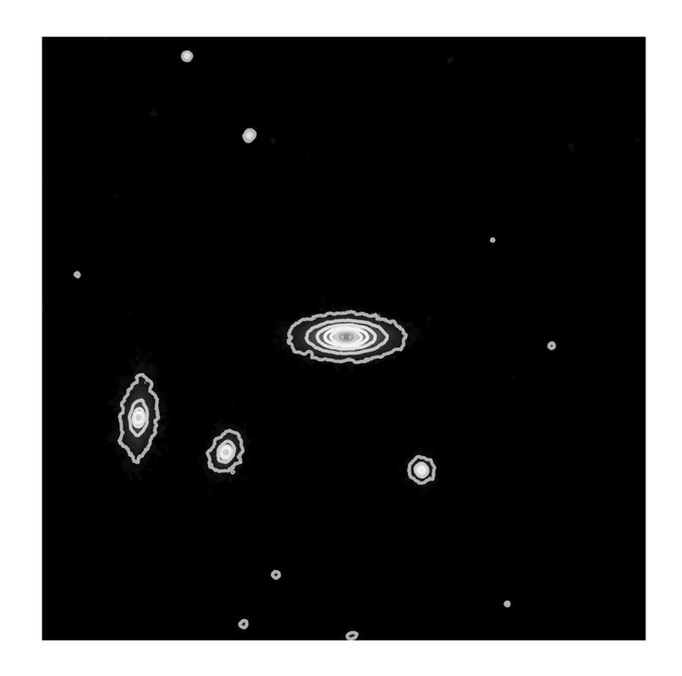
$$<I_{\text{input}}>=2.5\text{Median}_{m\times m}\circ I_{\text{input}}-1.5\text{Mean}_{m\times m}\circ I_{\text{input}}\,, \tag{8}$$

否则:

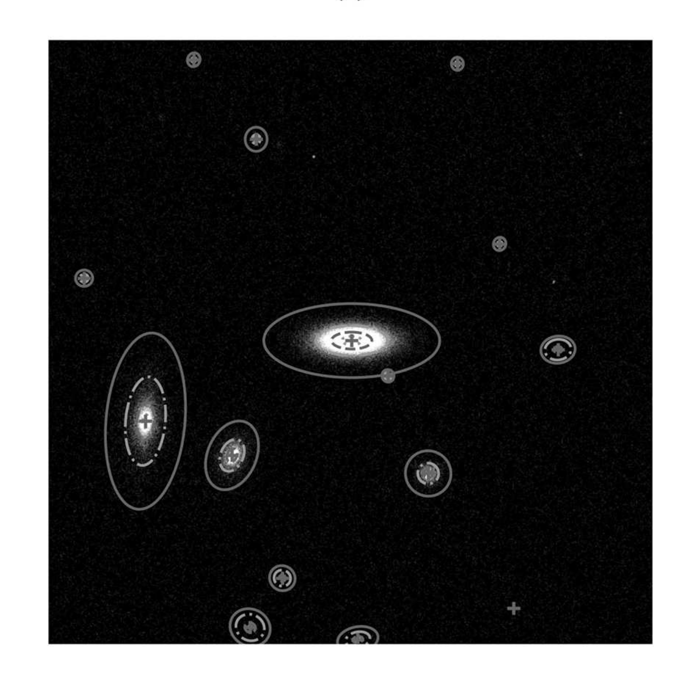
$$<I_{\text{input}}>=\text{Median}_{m\times m}\circ I_{\text{input}}\,, \tag{9}$$

滤波核尺寸控制了背景噪声模型构建的细致程度. 该滤波核尺寸m应大于典型点源而小于典型展源的线度.

除利用滤波方法获得图像均值外, 还可利用测光源列表的参数建模获得图像均值. 相比于滤波方法, 建模方法的优势是对展源中心部分的刻画更加细致. 但建模方法会忽略展源的非椭圆结构, 有着更高的时间成本并且测光误差可能会引起错误建模. 因此模块一般采用滤波方法获取像素均值, 仅对部分椭圆星系核心区采用建模方法进行均值估计. 滤波方法与建模方法获得背景噪声模型的示意图, 如图3所示.



(a)



(b)

图 3 滤波方法与建模拟合方法获得噪声模型的示意图. (a)滤波方法, (b)建模拟合方法.

Fig. 3 Noise model obtained by filtering and model fitting sketch map. (a) Filtering method, (b) model fitting method.

以椭圆星系为例, 滤波方法与建模方法在图像剖面的对比, 如图4所示.

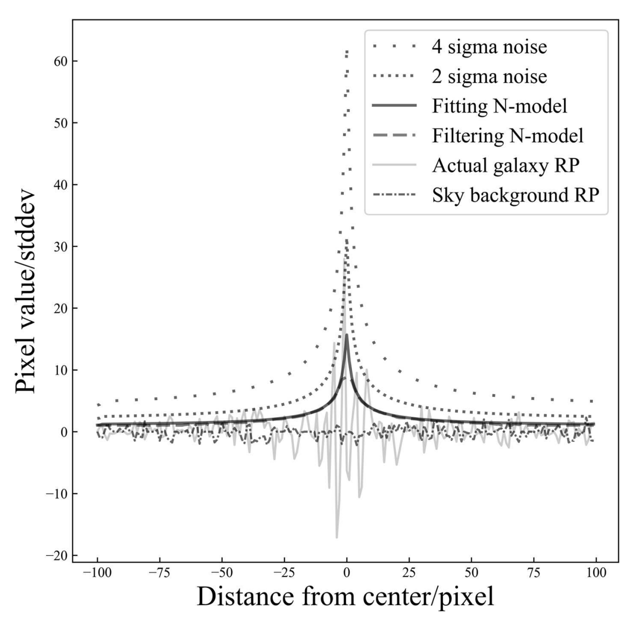


图 4　椭圆星系噪声建模方法对比(图例中"N-model"为噪声模型缩写; 图例中"RP"为residual profile缩写, 意为残差剖面; 为了方便后续流程对源亮度的描述, 在后续作图中将残差图像标准差$\sigma_{\rm std}$, 简计stddev, 作为表征亮度的坐标轴标准单位).

Fig. 4　Comparison of two methods for elliptical galaxy noise modeling (The"N-model" in the legend is an abbreviation for noise model; The"RP" in the legend is an abbreviation for residual profile; In convenience of describing the source brightness in the subsequent process, the standard deviation of residual image $\sigma_{\rm std}$, abbreviated as stddev, will be used as a standard unit for coordinate axis representing brightness in subsequent plotting).

为了更加准确地去除宇宙线等随机因素的影响, 还可以取输入图像与模板图像二者最小值滤波后的图像进行均值估计. 进一步定义$\Phi_{\rm sky\text{-}mean}$为最小值滤波图中没有天体区域的平均天光背景通量(该值作为统计量已在第一部分第二流程获得), 定义残差图像标准方差为:

$$\sigma^2_{\rm std}\ = 2\varepsilon\alpha\Phi_{\rm sky\text{-}mean}\ tg^2 + 2\varepsilon\sigma^2_{\rm const}\ . \qquad (10)$$

对于给定输入图像, 残差图像标准方差是常数.

3.3　残差图像暂现源判定

残差图像暂现源(变源)判定流程是CSST暂现源探测方法的核心, 也是创新之处. 其难点在于如何区分CSST残差图像中的暂现源与非科学数据假源. CSST残差图像中的非科学数据假源共有4类, 其中占绝大多数的是来自于星光背景噪声涨落产生的残影. 另外3类还包括少量无法与点源区分的宇宙线、在宇宙线去除阶段误去除的点源以及在宇宙线遮罩时因不良噪声插值产生的局部极值. 对于后两类假源, 可以通过对比原始图像信息以及宇宙线标记星表(星表1)来进行筛选. 因此, 真正影响残差图像暂现源判定的假源即噪声涨落假源与无法与点源区分的宇宙线. 本工作希望通过某种方法针对性筛选占比较高的噪声涨落假源, 残差图像暂现源判定的流程, 如图5所示.

下面来具体考察噪声涨落假源的形成原因. 若在一处星光背景较小的区域, 仅仅通过噪声涨落而得到一个能够通过源识别判据(亮度、面积、椭率、锐度等)的假源是几乎不可能的. 由噪声涨落产生的假源一定是在本身背景噪声较大且具有一定形态分布的区域(例如星系中的非椭圆结构、前景弥散展源等), 通过散粒噪声的涨落得到并通过所有源识别判据的. 这类假源的亮度通常略高于当地背景噪声亮度, 不会明显亮于当地背景. 因此, 本工作解决残差图像暂现源判定问题的基本思想是: 考察残差图像(削波后)测光源候选列表中每个候选源与其当地泊松噪声的偏离程度. 候选源亮度越接近泊松噪声背景, 则该候选源是噪声涨落假源的概率越大. 反之, 若候选源亮度远远偏离当地背景噪声预期, 则说明该候选源更有可能是真实暂现源而非由噪声涨落产生的假源.

在应用中, 首先将残差图像与当地背景噪声模板的对应像素值相除, 并将所得到的二维数组称为对比图, 即图5中S. 继而对对比图进行测光, 并设置阈值来筛选高于阈值的真实暂现源. 本工作将这种方法称为对比度测光法, 也将对残差图像的测光称为直接测光法. 若残差图像测光(直接测光)源候选列表中的候选源同时出现在对比度测光候选列表中, 则将该源判定为真实源, 否则将该候选源判定为噪声涨落假源. 以椭圆星系、旋涡星系及无天体天空背景区域为例的残差图像、噪声模板及对比图窗口像, 如图6所示.

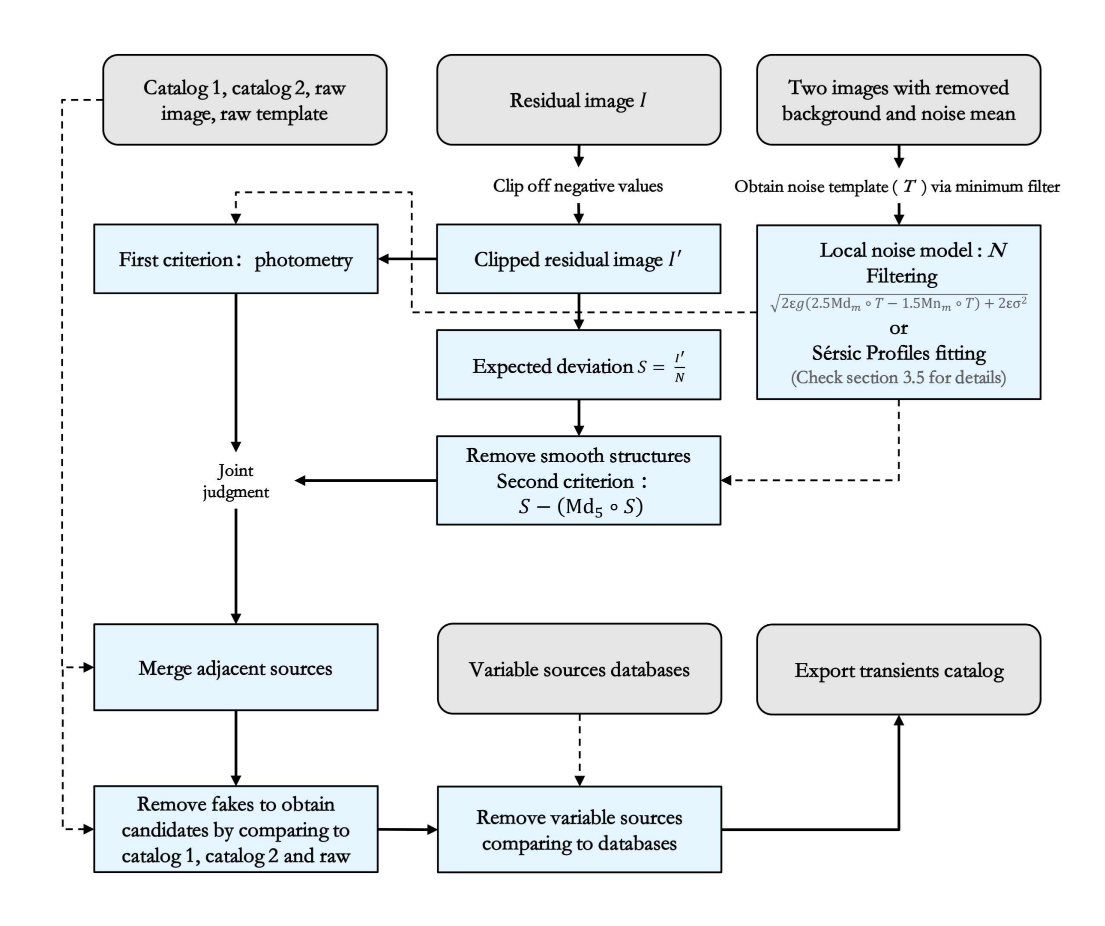


图 5 残差图探测流程(实线代表程序流向, 虚线代表数据流向; 简计$\mathrm{Mean}_{m\times m}$为$\mathrm{Mn}_m$, 简计$\mathrm{Median}_{m\times m}$为$\mathrm{Md}_m$; 定义残差图像为$I$, 削波后的残差图像为$I'$, 噪声模板为$T$, 噪声模型为$N$, 定义$I'$与$N$的比值为对比图, 记为$S$, 详见下文.)

Fig. 5 The process of residual image detection (While solid line represents program flow, dashed line represents data flow; $\mathrm{Mean}_{m\times m}$ is simplified as Mn_m, $\mathrm{Median}_{m\times m}$ is simplified as Md_m; Define residual image as I, clipped residual image as I', noise template as T, noise model as N, and define the ratio of I' to N as contrast graph, denoted as S, refer to the following text for details.)

判定流程首先对残差图像进行负流量削波. 流程中所使用的两种测光方法仍然基于DAOPHOT算法进行测光, 负流量信息不仅对于探测暂现源没有帮助还会对测光造成干扰. 因此仍需对残差图进行削波, 这一步同时去除了正在变暗的变源. 与此同时, 流程利用两张输入图像的最小值滤波作为噪声背景模板, 这样可以去除输入图像中包括残留宇宙线以及暂现源在内的所有随机因素, 可以获得较为“干净”的噪声背景模板. 接下来即利用该模板, 运用(7)式反推得到残差图像噪声模型. 对于对比度测光法, 由于常值噪声的存在, 不会出现削波残差图与使用滤波方法获得的噪声模型相除时分母为0的情况. 根据方法初步测试结果, 当待考察点源出现在背景噪声较小的区域时, 对比图的像素数值标准差变大, 测光过程容易受到残留平滑结构的影响. 而当点源出现在星系中使其背景噪声变大时, 则容易受到来自星系非椭圆结构的影响. 因此, 为了防止平滑结构在对比图测光时造成干扰, 流程在对比图测光前对其进行了额外的滤波以消除这些平滑结构. 将一幅图像减去$m\times m$中值滤波器与该图像的卷积, 即可抹除图像中一切大于$m\times m$的平滑结构. 此方法受到了L.A.Cosmic宇

宙线去除算法[21]的启发, 并针对CSST残差图像的特点加以改进. 最后, 在对残差削波图运用直接测光法与对比度测光法进行真实暂现源筛选后, 即得到残差图像暂现源候选列表.

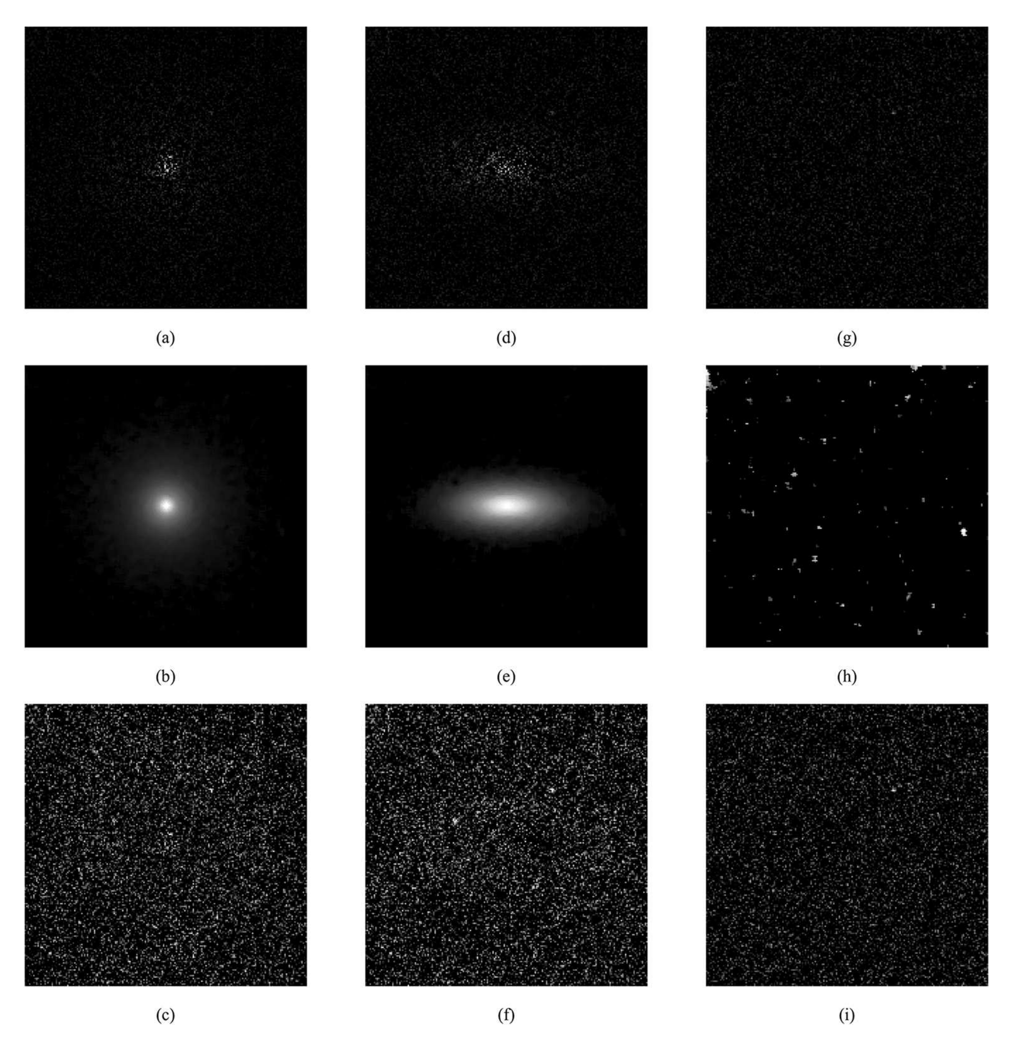


图 6　3种典型邻域的残差图窗口像、噪声模板窗口像及对比图窗口像. (a)椭圆星系邻域残差图像. (b)椭圆星系邻域噪声模板. (c)椭圆星系邻域对比图. (d)旋涡星系邻域残差图像. (e)旋涡星系邻域噪声模板. (f)旋涡星系邻域对比图. (g)天空背景残差图像. (h)天空背景噪声模板. (i)天空背景对比图.

Fig. 6　Window image of residual images, noise templates and contrast graph from three typical neighborhoods. (a) Residual image of elliptical galaxy neighborhoods. (b) Noise template of elliptical galaxy neighborhoods. (c) Contrast graph of elliptical galaxy neighborhoods. (d) Residual image of spiral galaxy neighborhoods. (e) Noise template of spiral galaxy neighborhoods. (f) Contrast graph of spiral galaxy neighborhoods. (g) Residual image of sky background. (h) Noise template of sky background. (i) Contrast graph of sky background.

3.4　测光方法阈值

在介绍如何确定测光方法阈值之前, 先明确探测率与准确率的概念. 在天文领域中不常用到统计学习领域查全率(Recall)与查准率(Precision)的概念, 与之对应的是探测率与准确率. 其中, 探测率的概念可以类比查全率, 即最终出现在暂现源星表中的源占测试输入源总数的比例. 还需要说明准确率

的概念, 可以类比于查准率, 即暂现源星表中真实输入暂现源占暂现源星表总数的比例.

在实际应用判定流程时, 应确定流程中各方法的最优参数以得到方法最佳性能. 直接测光法与对比度测光法各自的测光阈值是判定流程中最重要的参数. 工作首先在天空背景区域通过实验的方式确定方法阈值. 定义最小可分辨(点)源, 即使用1像素源点卷积1像素高斯展宽所形成的点源作为测试目标点源. 将点源亮度设为4倍天光背景标准差($\sigma_{\rm std}$, stddev). 在图7中展示了对目标区域使用两种方法不断交替降低阈值测光, 如图7 (a)所示, 以及所选阈值对目标点源探测率与准确率变化, 如图7 (b)所示.

可以看到, 有3种不同的阈值选择: 两种方法中仅一种识别了目标点源; 两种方法除了目标点源外均识别了其余非科学数据假源, 但两种方法和判定得到的仅仅只有目标点源; 以及第3种情况两种方法除了目标点源外还识别了许多假源, 并且两种方法的和判定源同样也包含了数个假源. 将这3种情况所对应的阈值称为: 保守阈值、和判定阈值与低阈值(作为参考). 为了尽可能识别更暗的暂现源, 和判定阈值中的最低值是方法希望采用的. 保守阈值、和判定阈值与低阈值对应图像示意图, 如图8所示. 示意图中所加入的测试目标点源位于图右上侧, 即图8 (a)与图8 (d)圈中所示点源.

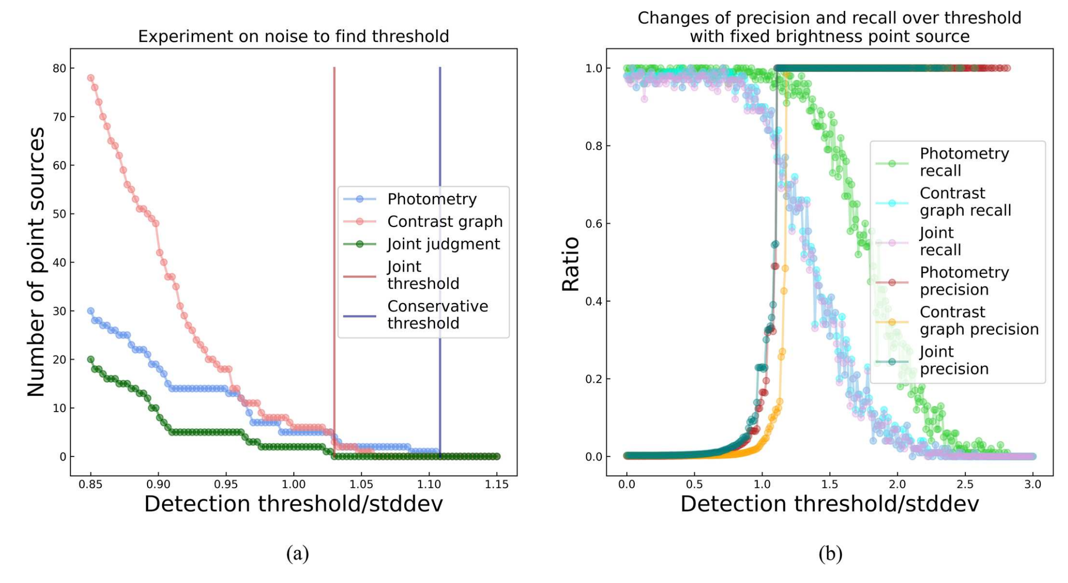


图 7　(a)不同阈值所对应的探测情况. (b)方法阈值对固定亮度点源探测的影响.

Fig. 7　(a) Different situations corresponding to different thresholds. (b) Influence of method threshold on fixed brightness point source detection.

接下来尝试改变目标点源的亮度以观察3种不同阈值下, 方法探测率与准确率的变化, 如图9所示.

可以看到, 若方法选取保守阈值, 虽然准确率可以一直保证为100%, 但两方法的探测率曲线相对于和判定阈值与低阈值的情况, 曲线中部整体右移, 即对同样亮度点源的探测率降低. 随着点源亮度从7倍天光背景标准差降低到接近背景噪声大小时, 3种阈值所对应的两方法探测率曲线差异, 均逐渐变大后变小趋同. 当点源亮度约2–3倍天光背景标准差时, 两方法出现最大差异. 对于低阈值情况, 完全无法保证方法探测准确率, 因此仅作参考. 当点源亮度大于天光背景噪声标准差时, 使用最小和判定阈值的判定方法可以做到“查到皆准”. 即当点源亮度大于0.97倍天光背景标准差时, 准确率一直为1. 只有在点源亮度小于0.97倍天光背景标准差

时, 才会出现准确率急剧下降的情况.

进一步, 将天空背景下最小和判定阈值的探测率曲线进行拟合. 测试目标点源添加在输入图像中, 并经过了之前的所有流程, 无法通过简单逻辑或理论推导去解析地求得曲线来源, 从而推断拟合曲线应该用什么形式的函数. 所以对于拟合函数形式如何选择的问题, 在当前阶段暂且直接通过曲线型状进行选择, 利用乙状曲线簇(Sigmoid函数)[10]来进行函数拟合:

$$S(x) = \frac{1}{1 + \mathrm{e}^{-(ax-b)}}\,, \tag{11}$$

其中a、b为控制曲线形状的两个参数. 拟合结果如图9 (d)所示.

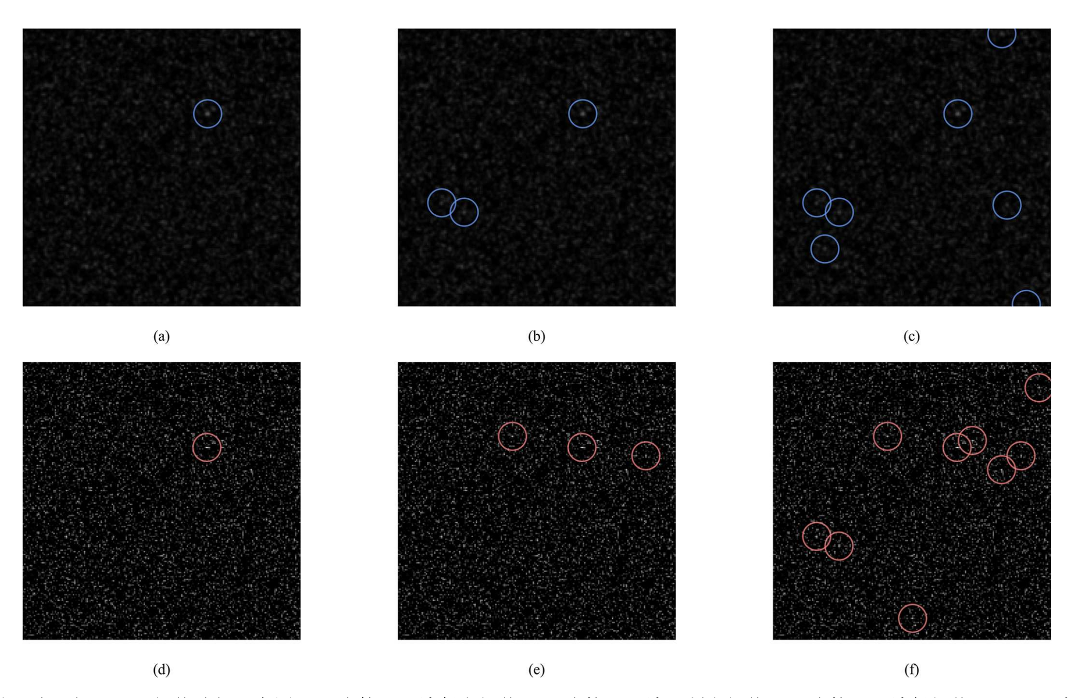


图 8 天空背景点源探测不同阈值选择示意图. (a)直接测光法保守阈值. (b)直接测光法和判定阈值. (c)直接测光法低阈值. (d)对比度测光法保守阈值. (e)对比度测光法和判定阈值. (f)对比度测光法低阈值.

Fig. 8 Different threshold selection for source detecting in sky background. (a) Conservative threshold for photometry. (b) Joint threshold for photometry. (c) Lower threshold for photometry. (d) Conservative threshold for contrast graph. (e) Joint threshold for contrast graph. (f) Lower threshold for contrast graph.

上述实验确定了应选取最小的和判定阈值作为方法阈值的原则. 天空背景有着统一的背景噪声水平, 因此可以使用最小和判定阈值作为单一阈值进行测光. 在实际情况中, 观测图像不可能没有任何背景天体. 若以点源为探测目标, 则观测图像中的星系展源则成为了测光背景, 对点源测光的背景噪声水平不再统一. 测光方法无法再继续使用单一阈值工作, 因为此时的孔径流量标准无法统一.

[10]查阅网址: https://mathworld.wolfram.com/SigmoidFunction.html

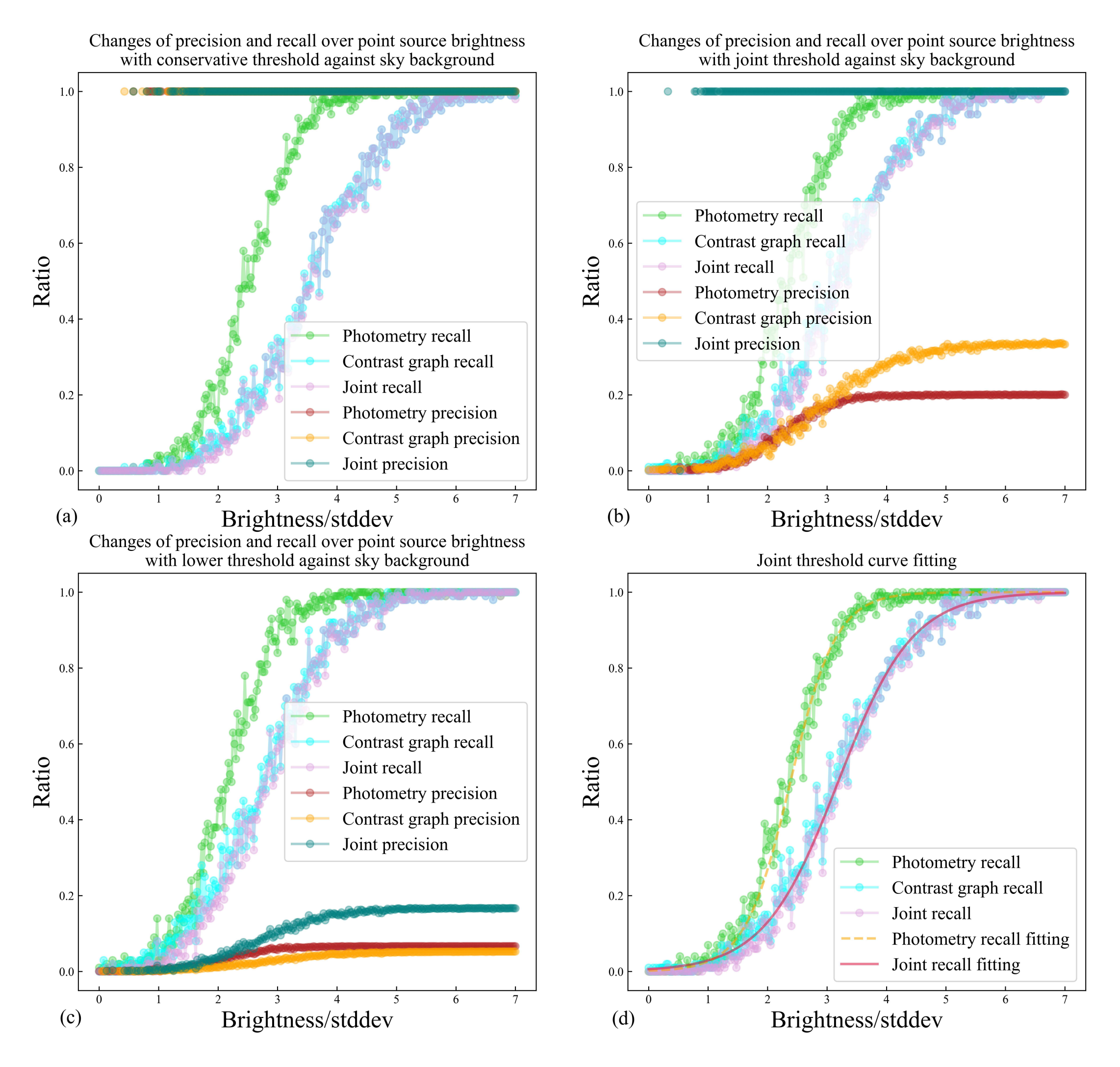
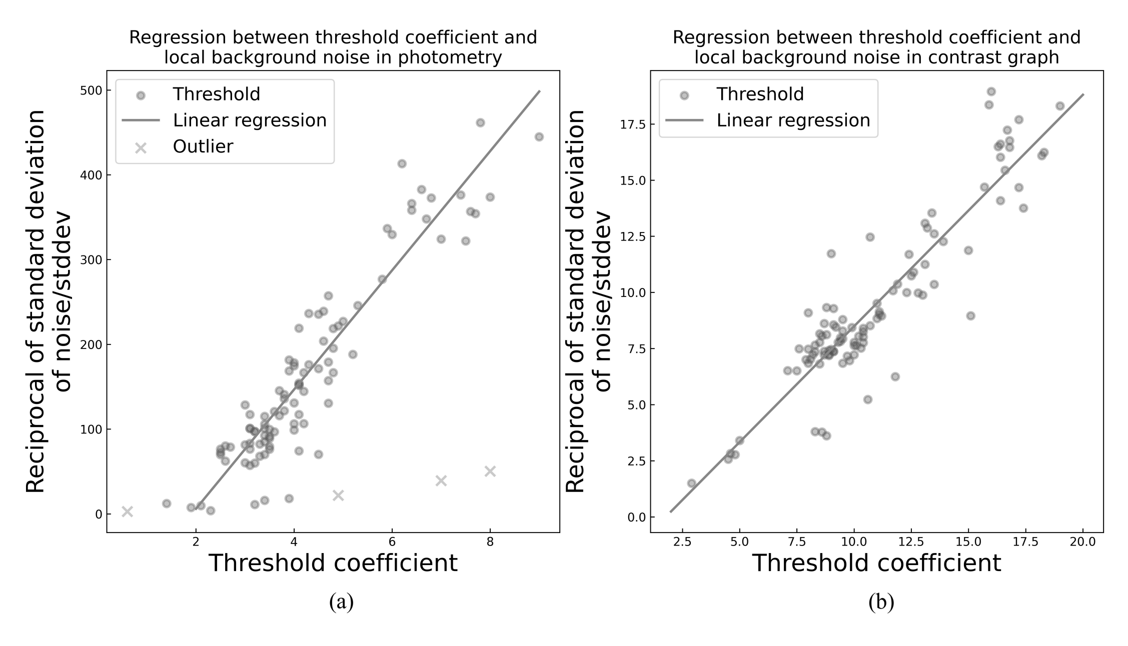


图 9 点源亮度的变化对不同阈值探测情况的影响以及和判定阈值查全率曲线拟合

Fig. 9 The influence of point source brightness on the detection of different thresholds with joint threshold recall curve fitting

因此, 在拥有连续变化背景亮度的区域, 工作拟采用渐变阈值法间接统一背景噪声水平来处理出现在展源邻域的暂现源. 其基本思想是: 希望找到背景噪声标准差与测光方法阈值之间的映射关系, 再将图像按照天空背景光度梯度分割为不同的等光度区域, 使得各个等光度区域能根据当地背景噪声确定方法阈值, 最终每个区域都能够根据两方法阈值确定和判定阈值, 从而间接实现背景噪声水平统一. 等光度区域的宽度应该大于图像中典型点源的线度.

基于此思想, 可以通过从小到大改变输入图像天光背景噪声大小, 按照前述方法寻找不同噪声水平与方法阈值之间的映射关系. 经实验发现, 直接测光法阈值系数与对比度测光法阈值系数以及二者的和判定阈值系数均与当地噪声标准差呈反比, 即与当地噪声标准差的倒数成正比. 直接测光法阈值系数及对比度测光法阈值系数与当地噪声标准差的关系如图10所示. 暂现源探测流程即可依此确定不同背景噪声水平所对应的测光阈值, 因此适用观测图像所有区域.



(a)　　(b)

图 10　直接测光法阈值系数及对比度测光法阈值系数与当地噪声标准差的关系

Fig. 10　Relationship between the threshold coefficients of photometry method and contrast graph with the standard deviation of local noise

图10也反映了图像噪声水平逐渐增大时, 噪声从泊松分布向正态分布的过渡. 泊松噪声具有“长尾”分布的特点, 更有可能在测光阈值较大的地方出现因噪声涨落而产生的假源, 因此两方法均呈现出了斜率为正的回归直线.

3.5　方法性能测试

在确定方法阈值后, 需要利用仿真数据对方法进行验证并测试暂现源探测模块的极限性能. 事实上, 大多数爆发现象均在星系中发生, 且爆发源几乎都为点源. 为了更贴近真实情况, 流程测试区域选择在星系邻域进行测试. 定义星系邻域为从星系中心沿光度梯度下降到亮度高于天空背景亮度10%等光度线所包围的区域. 测试星系选择CSST深场图像视场中星系的典型线度(约200像素)(例: 星系LEDA 25177, 红移$z = 0.2048$, 对应CCD线度约205 pixels; 星系3C 009, 红移$z = 2.0197$, 对应CCD线度约94 pixels)[11].

用于程序测试的星系均由仿真程序按Sérsic光度轮廓[22]额外加入非椭圆特征生成:

$$I[r(x,y)] = I_{\mathrm{e}} \exp\left\{-b_n\left[\left(\frac{r}{r_{\mathrm{e}}}\right)^{(1/n)} - 1\right]\right\}, \tag{12}$$

其中I是图像中任意点的亮度, r是该点相对于源中心的距离. r_{e}是半光半径, 意味着在r_{e}内包含了星系总光度流量的一半, $L\left(r \leqslant r_{\mathrm{e}}\right) = \frac{1}{2}L(r \leqslant \infty)$. $I_{\mathrm{e}} = I\left(r_{\mathrm{e}}\right)$是在半光半径处的亮度, 决定了轮廓的幅值. n是Sérsic指数, 代表了轮廓曲线的陡峭程度, 指数越大轮廓越向源中心点集中. b_n是对公式中半光半径归一化后的修正系数, 可以通过关系$\Gamma(2n) = 2\gamma(2n, b_n)$求得, 其中$\Gamma$与$\gamma$分别代表了伽马函数与下不完全伽马函数. 本工作使用指数$n = 4$的轮廓描述椭圆星系以及漩涡星系星系核, 使用$n = 1$即指数分布的轮廓去描述漩涡星系星系盘, 使用$n = 0.5$即高斯分布的轮廓去描述代表暂现源的点源, 也被用来描述星系团中椭圆形的团间发光.

[11] 查阅网址: http://ned.ipac.caltech.edu/byname

在性能测试中, 希望得到暂现源探测方法对不同信噪比(Signal to noise ratio, S/N)暂现源探测率与准确率的变化. 在固定背景亮度下改变暂现源亮度或在固定暂现源亮度情况下改变背景亮度, 均可以改变目标源信噪比. 因此, 在星系邻域中, 使用渐变阈值法, 针对固定亮度最小可分辨点源, 通过改变其与星系中心的距离(等光度区间)测试探测率与准确率的变化, 进而验证方法可行性. 具体测试使用半径约100像素的旋涡星系与椭圆星系邻域, 对不同距离固定亮度点源, 在同一等光度区间内进行100次探测测试, 得到了3种不同亮度最小可分辨点源探测率与准确率的变化曲线, 如图11所示.

经测试, 在星系邻域中对暂现源探测的准确率, 基本保持了当源亮度大于背景亮度时"查到皆准"的规律. 准确率与探测率曲线的震荡特性与模拟次数不足以消除模拟图像背景噪声及星系非椭圆结构的随机性有关. 我们进一步利用仿真程序生成100张相同星系参数的不同星系邻域输入图像, 在不同输入图像中分别独立重复地进行等光度区间点源探测实验, 结果如图12所示.

图12展示了对星系邻域独立重复实验探测率曲线的拟合. 从中可以看出, 在星系邻域进行实验所得到的探测率曲线, 与从天空背景映射而来的探测率曲线有些许不同. 相比之下, 星系邻域得到的探测率曲线中部整体左移.

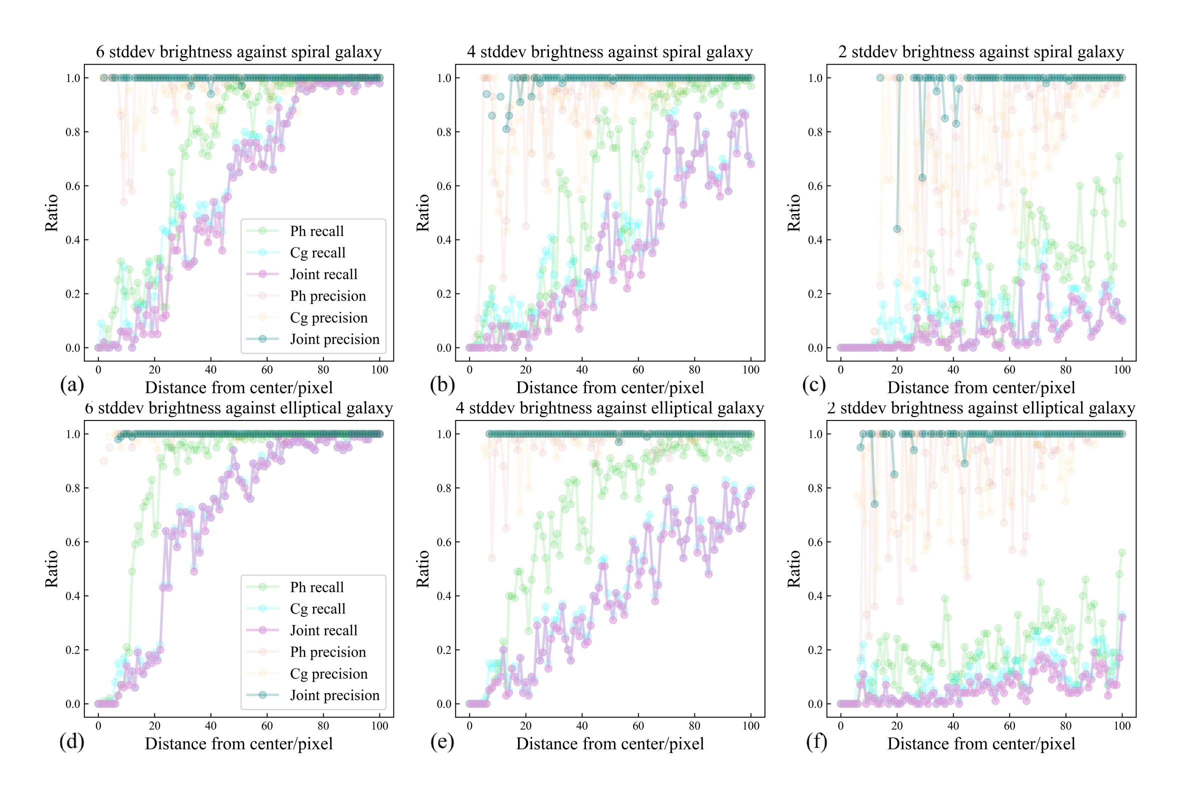


图 11 星系邻域方法探测率及准确率变化(图例及后文中, "Ph"是直接测光方法Photometry的缩写, "Cg"是对比度测光方法Contrast graph的缩写)

Fig. 11 Changes in recall and precision over galaxy neighborhood (In legend and subsequent text, "Ph" is the abbreviation for photometry method, and "Cg" is the abbreviation for contrast graph method)

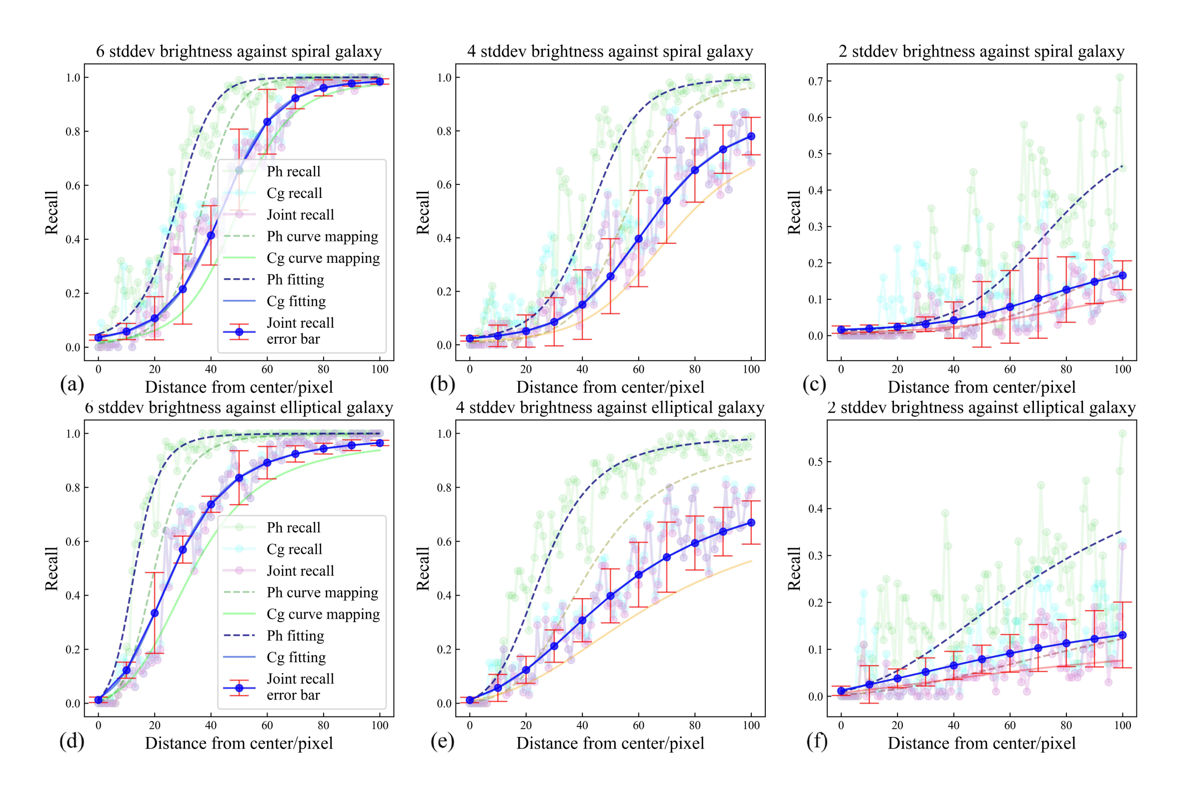


图 12 星系邻域独立重复实验探测率曲线拟合

Fig. 12 Fitting of recall curve over independent tests in galaxy neighborhood

这再次说明了星系邻域的噪声均值较大, 因此更加趋于正态分布, 而天空背景则相对趋于泊松分布. 考察重复实验的探测率曲线, 可以发现其相对于理论拟合曲线的误差呈现中间大两边小的规律. 理论上, 越靠近星系中心噪声标准差越大, 这将导致探测的不确定性更大. 但由于探测率的下降, 探测基数的减小加之其绝对数值接近0, 导致方差减小, 因此曲线在中部产生了最大误差.

最后, 图13展示了探测率随信噪比变化的关系曲线. 具体来看, 图13 (a)统一以点源信噪比为自变量, 对比展示了在星系邻域与天空背景中的探测率曲线. 星系邻域附近的噪声呈现正太分布, 而天空背景的噪声则接近泊松分布. 可以发现来自星系邻域与来自天空背景的曲线型状基本保持不变, 而星系邻域的曲线相比于天空背景曲线中部向左平移. 可见, 符合正态分布的噪声相对于符合泊松分布的噪声背景会产生更好的点源探测率. 这符合物理直觉, 因为相同均值的泊松分布所产生的“长尾”, 相对于正态分布更容易产生大的涨落. 这些因涨落产生的局部极值在经过滤波与测光后, 容易成为假源. 另一个可能的原因是, 当背景噪声较大时, 相同信噪比的暂现源绝对亮度数值更大, 与背景噪声的数值差值也更大. 在进行测光前滤波时, 相对于处在天空背景上的暂现源受影响更小. 后者与当地噪声的绝对差值更小, 因此相对更容易在测光源提取判据中受到影响, 可能因没有通过源提取判据而被误认为是噪声.

图13 (b)则展示了在200张不同输入图像的全图范围内, 分别加入100个暂现源进行实验所得到的探测率-信噪比曲线. 这些暂现源的亮度符合高于天光背景均值0–10倍标准差的均匀分布. 为了更加贴近真实情况, 暂现源的模拟产生位置更加侧重于星系邻域. 因此, 所得到的探测率曲线整体更加贴近由星系邻域反推所得到的探测率信噪比曲线.

误差棒也说明了几乎所有数据都夹在了天空背景与星系邻域探测率曲线之间. 可以看出, 星系邻域的流程性能测试所得到的信噪比曲线更加接近真实情况. 以上即完成了暂现源探测模块的性能测试, 并验证了CSST残差图像暂现源探测方法的可行性.

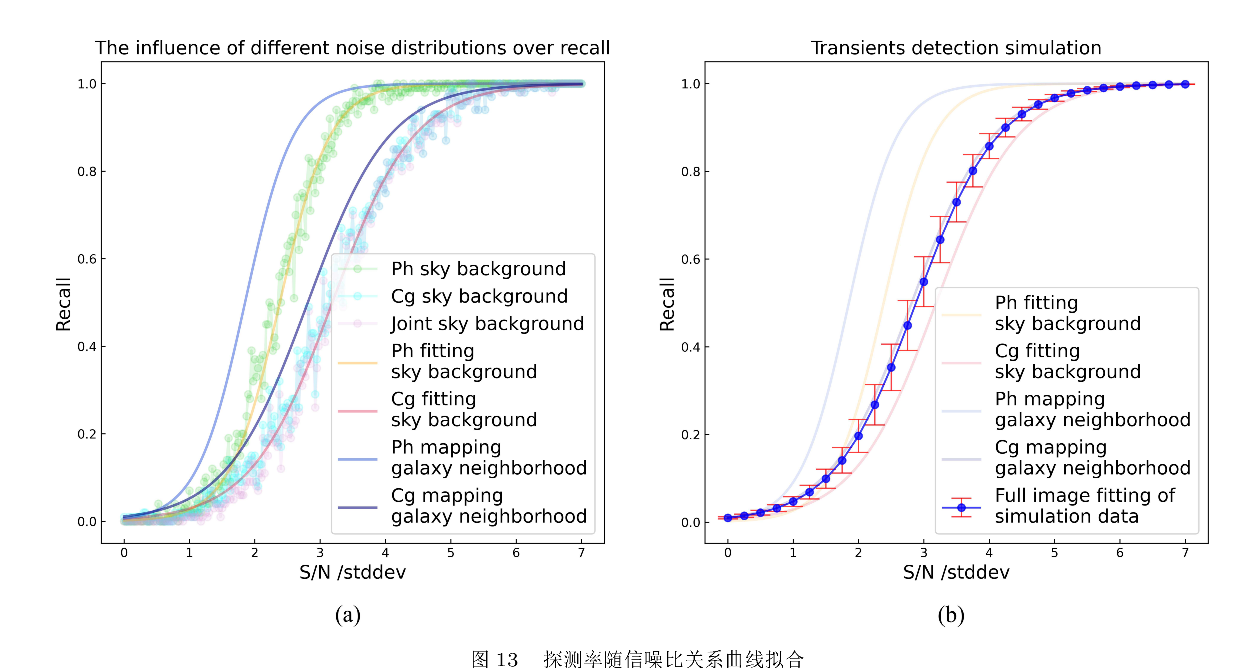


图 13 探测率随信噪比关系曲线拟合

Fig. 13 Curve fitting of recall to signal-to-noise ratio

4 总结与讨论

4.1 总结

暂现源探测模块在从输入图像到输出星表的全流程性能测试中, 使用200张由仿真程序生成的尺寸约9.2k×9.2k像素的不同主巡天观测图像, 总计约20000个目标暂现源以及约450000个天体. 测试最终定量给出了对目标暂现源探测准确率接近96%的情况下, 不同信噪比目标源探测率的变化图. 残差图像暂现源判定流程对于输入图像中暂现源的探测准确率为93.2%, 在弥补1.6%的误去除点源与1.1%的宇宙线不良去除占比后, 暂现源探测模块全流程对暂现源的探测准确率为95.9%, 探测错误率约4.1%. 在误识别的假源中, 宇宙线残留占比约37.2%, 噪声残影假源占比62.8%. 与平均探测正确率95.9%相对应的探测率, 符合Sigmoid函数, 其型状参数a、b如下:

$$\begin{aligned} \mathrm{Detectivity(S/N)} &= \frac{1}{1+\mathrm{e}^{-(a\times \mathrm{S/N}-b)}}\,, \\ a &= 1.601 \pm 0.019\,, \\ b &= 4.608 \pm 0.239\,, \end{aligned} \tag{13}$$

曲线型状参数a、b各自的边缘分布接近正态分布, 二者分布的标准差分别为0.019以及0.239. 相较于先前工作, 本工作利用仿真数据针对空间望远镜图像特征, 更加全面地定量给出了不同亮度暂现源所对应的探测率. 带有随机性的多轮独立性能测试也验证了模块中残差图像暂现源判定方法及全流程的可行性、独立性、泛化能力与稳定性. 在此需要再次说明的是, 工作中的各种测试结论均在点源尺寸为最小可分辨点源的情况下得到. 经过测试对比, 当点源展宽大于最小可分辨点源时, 所有测试结论一致优于最小可识别点源的情况. 因此, 测试结论相当于是模块的极限性能测试.

4.2 讨论

暂现源探测模块流程与方法具有良好的兼容性, 可以作为逻辑模型应用到其他有类似需求的巡天项目中. 考虑到暂现源探测模块在开发过程中均基于CSST观测策略与数据特点进行特异性调整优化. 因此, 如若希望在其他巡天项目中应用此模块, 仍需要针对该项目的巡天策略与观测数据特点对模块进行参数调整与优化. 同时, 在本工作中还存在一些有待完善的问题. 主要有以下3点.

第一, 仿真图像相对真实望远镜的FITS图像肯定还存在差距且存在没有考虑到的效应, 如饱和星芒、前景展源等; 现在的图像相对真实图像仍然显得太"干净"; 因此在望远镜实际运行获得真实图像后, 仍需对没有考虑到的仪器或物理效应进行针对性流程改进.

第二, 本工作的测光源列表分类特征来源于对源进行模型拟合的参数; 虽然根据这些模型参数已经获得了可执行的分类结果, 但这些参数都是高压缩度参数, 相对于原始图像损失了许多信息; 在传统机器学习方法中是否存在分类效果更好的人工给定特征以及算法来区分非科学数据, 是将来要进行研究的问题; 特别是在解决假源模型制作问题后, 完全可以采用深度学习方法在不损失任何图像信息的情况下进行监督学习分类训练.

第三, 在本工作中已经假设, 经过上游流水线预处理的两张暂现源探测模块输入图像处于完全对齐状态; 这要求高精度的特征源目标定位与重采样, 属于理想情况; 在实际观测中可能会出现图像边缘位置不能完美对齐的情况, 错开的展源残影在残差图上会影响残差图暂现源探测的进行; 因此, 在后续工作中一定需要根据真实图像重采样对齐情况针对错位对流程进行改进.

在此需要补充说明的是, 暂现源探测模块中"对比数据库去除变源"部分的变源数据库来自CSST数据流水线的其他模块, 旨在通过数据库信息比对将正在变亮的变源从暂现源列表中剔除. 在后续CSST数据处理流水线程序联调中将对该外部接口进行测试, CSST数据处理流水线可以保证该接口的正常输入.

致谢 感谢审稿人提出的宝贵建议, 使得文章质量有了显著提高.

参 考 文 献

[1] Li S Y, Li Y L, Zhang T M, et al. SCPMA, 2023, 66: 1869
[2] Infante-Sainz R, Trujillo I, Román J. MNRAS, 2020, 491: 5317
[3] 詹虎. 科学通报, 2021, 66: 1290
[4] Brink H, Richards J W, Poznanski D, et al. MNRAS, 2013, 435: 1047
[5] du Buisson L, Sivanandam N, Bassett B A, et al. MNRAS, 2015, 454: 2026
[6] Goldstein D A, D'Andrea C B, Fischer J A, et al. AJ, 2015, 150: 1538
[7] Jia P, Zhao Y, Xue G, et al. AJ, 2019, 157: 250
[8] Alard C, Lupton R H. ApJ, 1998, 503: 1538
[9] Blanter Y M, Buttiker M. PhR, 2000, 336: 1370
[10] Wu J J, Zhang Z C, Tian Y, et al. A Method of CCD Noise Removal in Digital Images. 2nd IEEE Conference on Industrial Electronics and Applications. Harbin, China: IEEE, 2007, 55: 2571
[11] Stetson P B. PASP, 1987, 99: 191
[12] Fawcett T. PaReL, 2006, 27: 861
[13] Niculescu M A, Caruana R. Predicting Good Probabilities with Supervised Learning. Machine Learning, Proceedings of the Twenty-Second International Conference (ICML 2005). Bonn, Germany: ICML, 2005, 4: 625
[14] Christopher B M. JEI, 2006, 16: 140
[15] Geoffrey H E. AI, 1989, 40: 185
[16] Breiman L. MLS&T, 2001, 45: 5
[17] Freund Y, Schapire R. JCSS, 1997, 55: 119
[18] Friedman J H. AnSta, 2001, 29: 1189
[19] Chen T, Guestrin C. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. California, San Francisco, USA: ACM, 2016, 1: 785
[20] Rasmussen C E, Williams C K I. Gaussian Processes for Machine Learning. Massachusetts: MIT Press, 2005: 63
[21] van Dokkum Pieter G. PASP, 2001, 113: 1420
[22] Sérsic J L. Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 1963, 6: 41

Transient Detection Method Based on CSST

ZHAO Jie-ming[1] WANG Xiao-feng[1,2] LI Guo-liang[3] ZHANG Tian-meng[4]
(1 Physics Department, Tsinghua University, Beijing 100084)
(2 Beijing Planetarium, Beijing 100044)
(3 Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210023)
(4 National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012)

Abstract China Space Survey Telescope (CSST) is a 2-meter aperture optical space telescope planned by the China Manned Space Program, with the primary focus on large-scale surveys, which combines high spatial resolution and a large field of view. Transients are one of the important scientific objectives of CSST. As a space telescope, the image noise composition of CSST is significantly different from that of ground-based telescopes. Therefore, it is unsuitable to directly apply the transient detection scheme of previous ground-based survey project, and it is necessary to develop an independent detection method for CSST. The CSST transient detection method is developed and validated based on the CSST simulation data, aiming to identify transients in CSST main survey data as accurately as possible. The method is based on Difference Image Analysis (DIA) method, which firstly obtains the residual images corresponding to the observed image, and then uses filtering method to obtain the local Poisson noise distribution of the residual image. The method then uses a changing threshold to indirectly unify the image noise level, and performs photometry and source extraction on the residual image to obtain a list of candidate sources. Next, the real sources are filtered by judging the deviation of the candidate sources from local Poisson noise. After combining the photometric classification information of the observed image and the information of original image, the fake sources in the candidate source list are further eliminated. The final output is a transient list. The method is able to filter transients in the observed images with an average accuracy of 95.9% in a multi-round simulation test with a total of 20000 transients. Compared with the previous work, the test also quantifies the detection rates of transients with different brightness more comprehensively. The conclusions of the test validate the feasibility, generalisation and stability of the CSST transient detection method. The method also provides a theoretical and programming basis for transient detection mission of space telescopes.

Key words methods: data analysis, techniques: image processing, telescopes, surveys